

# Rate of correct detection of periodic signal with the *Gaia* satellite

L. Eyer<sup>1,2\*</sup> and F. Mignard<sup>3</sup>

<sup>1</sup>Geneva Observatory, CH-1290 Sauverny, Switzerland

<sup>2</sup>Princeton University Observatory, Princeton, NJ 08544, USA

<sup>3</sup>Observatoire de la côte d'azur, CNRS, BP 4229, 06304, Nice, France

Accepted 2005 May 24. Received 2005 May 11; in original form 2004 June 14

## ABSTRACT

The *Gaia* satellite was selected as a cornerstone mission of the European Space Agency (ESA) in 2000 October and confirmed in 2002 with a current target launch date of 2011. The *Gaia* mission will gather on the same observational principles as *Hipparcos* detailed astrometric, photometric and spectroscopic properties of about one billion sources brighter than  $V = 20$  mag. The nature of the measured objects ranges from near-Earth objects (NEOs) to gamma-ray burst afterglows, and encompasses virtually any kind of stars in our Galaxy. *Gaia* will provide multicolour (in about 20 passbands extending over the visible range) photometry with typically 250 observations distributed over 40 well-separated epochs during the 5-yr mission. The multi-epoch nature of the project will permit the detection and analysis of variable sources whose number is currently estimated in the range of several tens of millions, among the detectable sources.

In this paper, we assess the performances of *Gaia* in analysing photometric periodic phenomena. We first quickly present the overall observational principle before discussing the implication of the scanning law in the time-sampling. Then from extensive simulations one assesses the performances in the recovery of periodic signals as a function of the period, signal-to-noise ratio and position on the sky for simple sinusoidal variability.

**Key words:** methods: numerical – techniques: miscellaneous – surveys – space vehicles: instruments – stars: variables: other.

## 1 INTRODUCTION

The very successful European Space Agency (ESA) mission *Hipparcos* was the first modern multi-epoch all-sky space survey. Although primarily designed for astrometry, the full-light (supplemented by two chromatic bands of lesser accuracy) photometry proved to be a significant addition to the initial goal. About 10 per cent of the objects in the final catalogue have information about variability, 7 per cent are listed in the periodic and unsolved catalogues and 3 per cent (i.e. 3794) have been published with their light curves. A refined analysis of the occurrence of variability through the whole Hertzsprung–Russell (HR) diagram was subsequently achieved (Eyer & Grenon 1997).

The *Gaia* mission, a cornerstone of the ESA science programme, will dramatically improve this analysis in several respects: (i) the photometric precision will be much better, in the range of mmag for single observations (compared to 0.01 for *Hipparcos*); (ii) the number of objects will be increased by a factor of nearly 10 000; (iii) multiband photometry (15–20 photometric channels in addition to the full light) will be available and will permit the identification

of the physical processes responsible for the variable phenomena. In addition, there will be an instrument for low-resolution spectroscopy. The mission selected in 2000 and confirmed in 2002 with a modified design is scheduled for a launch in mid-2011 for a 5-yr sky survey.

Scientific preparation for the mission involves the participation of some 15 working groups sharing responsibility for the simulation, the data processing, the science modelling and the instrument optimization. Extensive simulation programs have been developed feeding all the groups with input data required to test the processing. In particular, the scanning parameters are now fixed and constitute the basic tool to determine the time-series. The algorithm used in this paper is based on the analytical algorithm built and implemented by one of us (Mignard 2001) and used thoroughly by the *Gaia* community.

Regarding the photometry, as of mid-2005, the choice of passbands and the parameters of the scanning law are now fixed. The relative positions of the fields of view (FOVs) may again change before the design is frozen by the prime contractor next year. We have used the design of the study phase of mid-2003. The measurement errors have also been assessed, and for the full-light magnitude (the so-called  $G$  magnitude) can be found in Jordi (2003). It is better than 0.002 mag for a single observation (one transit in the astrometric

\*E-mail: laurent.eyer@obs.unige.ch

field) of stars brighter than  $G = 15$  and of the order of 0.01 mag for the faintest stars ( $G \sim 20$ ) detectable with *Gaia*. This gives a rough idea of the wealth of science information recoverable on the variable stars, virtually over any spectral type or luminosity class.

The large-scale photometric survey will have significant value for stellar astrophysics, supplying basic stellar parameters (effective temperatures, metallicity and abundances, gravities), together with a virtually unbiased sample of variable stars of nearly all types, such as eclipsing binaries and pulsating stars with periods ranging from hours to several hundred days. Much shorter periods could be achieved also from the exploitation of the light recorded during the passage of the image on individual CCDs of a FOV. This particular and very specific sampling is not addressed in this paper. Variability search methods will be systematically applied to every photometric time-series, before the period search. Estimated number of variable stars to be detected as such is highly uncertain, but crude estimates suggest some  $20 \times 10^6$  in total, half of them being periodic variables (Eyer & Cuypers 2000).

## 2 THE OBSERVING INSTRUMENT

Although most technical details can be found on the *Gaia* documentation (e.g. Perryman et al. 2001), there are two items particularly relevant to the study of the variable stars, as they impact directly on the time-sampling. The first is the way the sky is scanned, allowing one to observe a given region of the sky at about 30–50 different times during the 5-yr mission, with an average return in the same direction every 6 weeks. The second is concerned with the repeated measurements over much smaller time-intervals, typically between 1 and 10 h. This sampling is determined by the relative positions of the different FOVs, through which photometric measurements are carried out. These two aspects of the design are discussed in the following.

The instrumental parameters used in this paper refer to the current baseline of the design at the time of writing. This is subject to a possible change during the study phase, without major impact on the main results of this paper, regarding the determination of the period of the variables, all the more because the simulation is primarily based on the data coming from the astrometric FOVs.

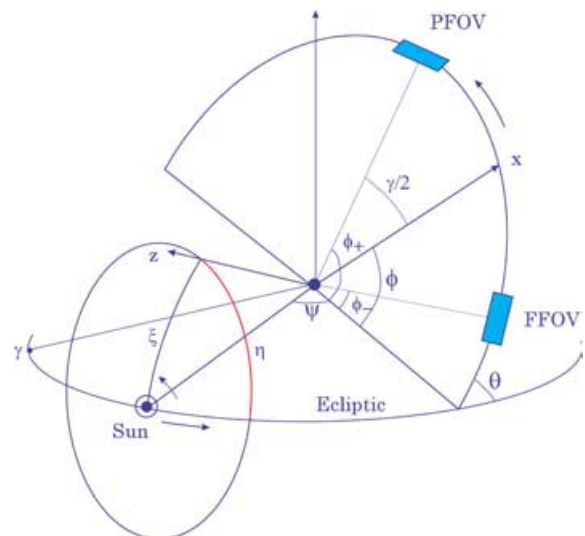
## 2.1 The scanning law

Measurements with *Gaia* are conducted with a scanning satellite with two widely separated astrometric FOVs, similar in its principle to the solution adopted for *Hipparcos*. This proved to be a very efficient and reliable solution for global astrometry.

The sky is scanned continuously following a pre-defined pattern, with the satellite spin axis kept at a fixed angle of  $50^\circ$  from the Sun (Fig. 1) and nominally perpendicular to each viewing direction. A sun-centred precession motion of the spin axis in  $\sim 70$  d allows a rather uniform coverage of the full sky every 6 months, yielding repeated observations of every region of the sky. The scan period is 6 h amounting to a displacement rate of the stellar images on the CCDs of  $60 \text{ arcsec s}^{-1}$ .

## 2.2 The photometric FOVs

The *Gaia* payload is fitted with two main telescopes with an aperture of  $1.4 \times 0.5$  m<sup>2</sup>, primarily dedicated to astrometry and full-light photometry. On each field, several columns of CCDs are covered with broad-band filters. This broad-band photometric (BBP) field provides multicolour, multi-epoch photometric measurements for



**Figure 1.** The nominal scanning law of *Gaia*. The spacecraft rotates around the z-axis in 6 h so that the FOVs sweep approximately a circle on the sky. The z-axis is constrained to move on a sun-centred cone of  $50^\circ$  half-aperture with a period of 70 d, forcing the plane of scan to sway back and forth with inclinations between  $40^\circ$  and  $140^\circ$ . The axis of the cone follows the yearly solar motion.

each object observed in the astrometric field, for chromatic correction and astrophysical purposes. Four or five photometric bands should be implemented within each instrument with an integration time of  $\sim 3.3$  s per CCD. A third telescope with a square entrance pupil of  $0.5 \times 0.5$  m<sup>2</sup> collects starlight for the spectrometer and the medium-band photometer (MBP).

### 3 SEQUENCES OF OBSERVATION

Because of the combination of the quick spin motion over 6 h and that of precession over 70 d coupled with the annual solar motion, the sampling of the epochs at which a particular object is measured is very peculiar and quite irregular. However, irregular does not mean random, and there are some basic recurring patterns that apply to each sequence. One must consider separately the succession of observations over a period less than  $\sim 1\text{--}2$  d due to the repeated passages over few consecutive rotations of the satellite, and are primarily governed by the spin and the focal plane layout, and the return in a particular direction after several weeks without observations, controlled by the precession motion. Therefore, a complete sequence divides itself naturally in succession of epochs, widely separated in time, in which there are a small number (between  $\sim 1$  and 10) of observations in the astrometric or photometric FOVs. For a regular sampling, the ability to carry out accurate frequency analyses of variable stars depends on the succession of repeated observations over short time-scales, that is, on sequences of successive revolutions separated by returns of the viewing direction to the same area of the sky. However, for a purely random sampling one can obtain alias-free results even for frequencies much higher than the inverse of the smallest intervals between observations, and one should not worry too much about the time-gaps between successive observations. However, *Gaia* sampling is obviously not perfectly regular, but also it falls short from a purely random sampling. This is something intermediate with regular patches over short time-scales repeating with much irregularity over longer time-scales. In the end,

**Table 1.** Fundamental sampling intervals following from the *Gaia* scanning law and from the location of the viewing directions. Transitions from an initial field to a final field are considered with the interval given in days. PFOV: preceding astrometric field of view, FFOV: following astrometric field of view, MBP: medium-band photometer field of view.

Initial field	Final field	Interval (h)
PFOV	FFOV	1.78
PFOV	MBP	5.38
PFOV	PFOV	6.00
FVOV	MBP	3.60
FFOV	PFOV	4.22
FFOV	FFOV	6.00
MBP	PFOV	0.62
MBP	FFOV	2.40
MBP	MBP	6.00

the combination of the two interleaved samplings proved important to achieve an ambiguous and accurate frequency retrieval.

The variety of short-term intervals likely to occur within a sequence of observations is given in Table 1. The highest temporal resolution is obtained by combining successive crossings in the MBP field with a photometric measurement obtained in the astrometric preceding field, leading to a minimum interval of 38 min.

### 3.1 Short-term succession

#### 3.1.1 Astrometric fields

Let a star be observed in the preceding astrometric field, at some reference time  $t = 0$ . The star is located somewhere in the FOV of width  $0.66$  in the cross-scan direction. We use a local rectangular frame with  $x$ -axis along the scan direction and  $y$ -axis along the transverse direction. Take the star image at the middle of the field for this reference time, that is,  $y = 0$ . Six hours later, this viewing direction is back in nearly the same region of the sky. However, the pole of the scan circle has moved by nearly  $1^\circ$  on the sky, and the ordinate of the star in the field can now be as large as  $1^\circ$ . In this case, the star is no longer in the FOV and cannot be re-observed. On the other hand, if the star happens to lie very close to one of the two stationary points of the scan circle (over a period of few days the scan of the sky can be described by the slow motion of a great circle, implying that there are two nodes on this circle that are stationary) then the star has virtually not moved on the field and will be observed again after one revolution. This situation could recur during four or five rotations of the satellite, supplying as many observations in the preceding FOV.

Consider now the transition between the preceding and following astrometric FOVs. They are separated by  $106^\circ$ , meaning that it takes 106 min to rotate from one field to the next. During this interval the apparent transverse motion of the star with respect to the scan circle is at most of  $0.3$ . Therefore, a star observed at a random position in the preceding field will be observed in the following field with a  $\sim 0.5$  probability. The average transverse motion being only  $2/\pi \times 0.3 = 0.19$ , this makes the odds more favourable for the observations to be grouped at least by pair. With smaller transverse motion (star closer to the instantaneous node) one may have four or five successive pairs of observations, each separated by 0.25 d. During these crossings, there are photometric measurements in full light, and in the five broad-band filters as well.

We can do a simple probabilistic modelling to assess the probability that an observation in a particular field is isolated (sequence of length 1) or followed by at least one observation in the next available FOV. It is reasonable to assume that the first observation of a sequence may occur at any ordinate in the field (this is a crude approximation, as the first observation of a sequence is more likely to occur near the upper or lower edge of the field). The chance of having the star observable in the next field depends on the transverse velocity and on the time-interval between the two fields. Let  $V_t$  be the transverse velocity and  $\Delta t$  the time-interval. One has  $\Delta t = 106$  min from preceding FOV (PFOV) to following FOV (FFOV), 254 mn from FFOV to PFOV and  $360 \text{ mn} = 0.25 \text{ d}$  for one revolution. The transverse velocity depends very simply on the longitude,  $\phi$ , of the optical axis on the instantaneous scan circle, and takes the form  $V_t = V_{\max} \sin(\phi + \alpha)$  with  $V_{\max} = 0.170h^{-1}$ , and where  $\phi$  is a fast-moving angle (period of 6 h) and the phase  $\alpha$  is a slow-varying variable with a typical time-scale of 70 d. We call  $X$  the random variable  $X = 1$  if the star is not re-observed after the first crossing at the start of a new epoch, and  $X = 0$  for the opposite outcome. If  $h$  is the transverse size of the field, it is easy to establish (see the Appendix) by conditioning the random variable to the transverse velocity and computing the expectation  $E(E(X | V_t))$  that

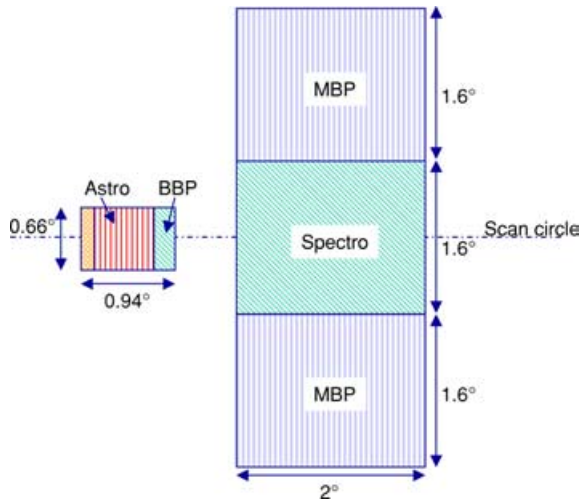
$$P(X = 1) = \begin{cases} \frac{2}{\pi} \frac{1}{u} & \text{if } u \geq 1 \\ \frac{2}{\pi} \frac{1}{u} [1 - (1 - u^2)^{1/2}] + \frac{\pi - 2 \arcsin(u)}{\pi} & \text{if } u \leq 1, \end{cases} \quad (1)$$

where  $u = h/V_{\max} \Delta t$ . When  $u > 1$  the transverse motion over  $\Delta t$  is always less than the width of the FOV, while it can be larger otherwise when  $u < 1$ . For the transitions between the preceding and the following FOV, this yields a probability  $P = 0.71$  that more than one observations occur, and 0.32 for the transition between the following and preceding fields, with the same assumption that the observation in the first field is the first of a sequence.

Table 2 gives the actual frequency of the different length of the sequence based on a simulation over 10 000 randomly distributed stars in the sky. The positions of these stars have been compared with

**Table 2.** Frequency distribution of the number of observations in continuous sequences. In the second and third columns only transits in the astrometric FOV are considered (i.e. in the BBP filters), while the last column includes also the MBP filters. The table gives the frequency of sequences of different lengths, accordingly the first field in the sequence is the preceding or following astrometric FOV (columns 2 and 3), while for the MBP, due to the size and position of the detectors, all sequences start with few consecutive observations in MBP, followed by transits including only BBP and MBP, before additional crossings of the MBP.

Length	Sequence starts with		
	PFOV	FFOV	MBP
1	0.27	0.63	0.00
2	0.60	0.20	0.00
3	0.05	0.11	0.06
4	0.04	0.02	0.13
5	0.01	0.02	0.33
6	0.01	0.00	0.22
7	0.00	0.00	0.07
>7	0.01	0.01	0.19



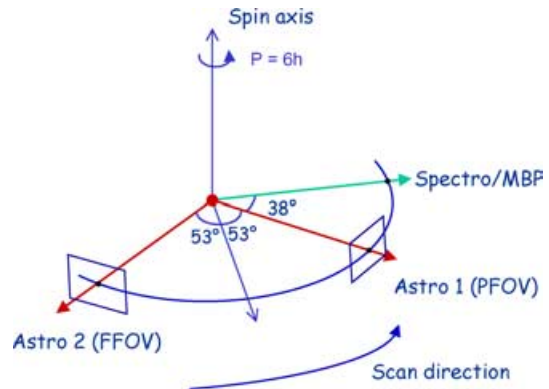
**Figure 2.** Relative size of the FOVs of *Gaia*. Many more field crossings arise in the medium-band photometer (MBP) than in the broad-band photometers (BBP) in the astrometric fields, as a result of the vertical extension.

the viewing directions over 5 yr, thus producing for each star the time-sampling in each FOV. The termination criterion for a continuous sequence of observations is activated when the interval between two successive observations is larger than few revolution periods of the satellite. A gap of three or four revolutions ( $\sim 1$  d) may happen in a short time-sequence in the passage from MBP transits to BBP transits as a result of the relative location of their respective detectors as shown in Fig. 2. Through trial and error we have found that a time-gap between two successive observations larger than 2.5 d was a good signature to isolate short observing sequences from very different observing epochs. This means essentially that when one has a gap of 2.5 d without observation, the actual gap will in practice be much larger ( $\sim 20$ – $30$  d), with few exceptions well accounted for by the idiosyncrasies of the scanning law.

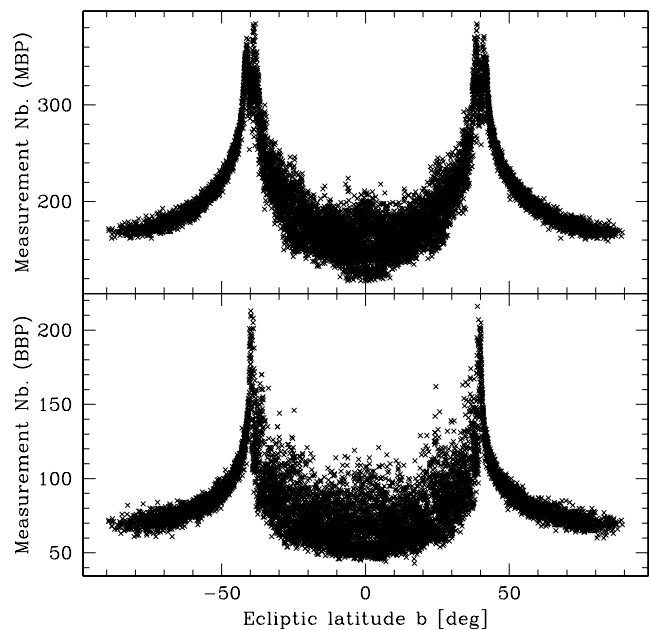
In the first two columns of Table 2, only the observations in the astrometric fields have been monitored and sequences starting with the preceding FOV (sequences like P, PF, PFP, PFPF ...) or following FOV (sequences like F, FP, FPF, FFPF, ...) are counted separately. They are not symmetrical, because the occurrence of a sequence PFOV  $\rightarrow$  FFOV is more likely than the opposite sequence. The typical sequence comprises two observations (PF) in the first group, while sequences of unit length (F) are the most common in the second group. The average length of the sequences is 2.1 in the first case and 1.6 in the second. While the number of long sequences is small (less than 1 per cent of the sequences longer than 10 observations), the length of the sequences can be very large and can reach 20 or 40 consecutive observations, meaning that the star is observed at regular interval for up to 5 d.

### 3.1.2 Astrometric and spectroscopic fields

We now introduce the field with the medium-band filters (MBP FOVs). It is located ahead of the preceding field, at  $38^\circ$  as depicted in Fig. 3 and comprises two subfields symmetrically placed with respect to the scan direction as shown in Fig. 2. Its vertical extension is much larger than the astrometric fields, meaning that the probability that a star is observed in this photometric field is higher than in the astrometric FOVs. The average number of observations is just above 200 in the MBP over the 5-yr mission.



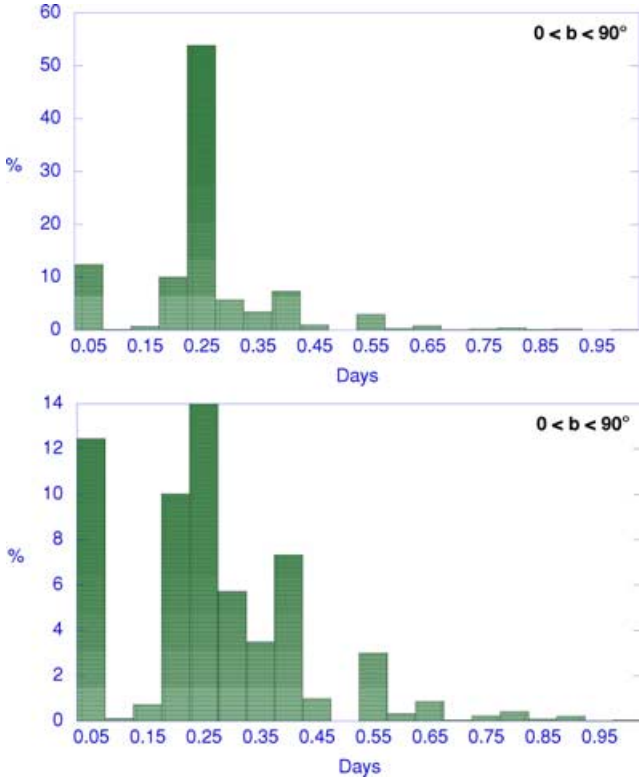
**Figure 3.** Relative positions of the viewing directions along the scan circle. PFOV (resp. FFOV) = preceding (resp. following) FOV.



**Figure 4.** Number of measurements in MBP (top) and BBP (bottom) FOVs as function of ecliptic latitude  $b$ . The mean number of observations is 200 for the MBP and 80 for the BBP.

The distribution with the ecliptic latitude (Fig. 4) displays the same symmetrical behaviour as for the observations with the broad-band filters.

The transverse motion over one revolution is always smaller than the vertical extension of any of the two subfields, which implies that observations take place over at least two successive revolutions of the satellite. In fact, the average length of a sequence consists of six to seven consecutive observations within less than a day, with measurements in every bandpass. In this case, all the sequences but a handful start with an observation in the MBP field, and a typical sequence is MMPFMM, meaning two MBP measurements followed by observations in the preceding and following astrometric fields and then again two in the MBP field. The probability distribution of the sequence lengths is given in the rightmost column of Table 2. Very long sequences of nearly 100 uninterrupted consecutive observations over  $\sim 10$  d may happen exceptionally.



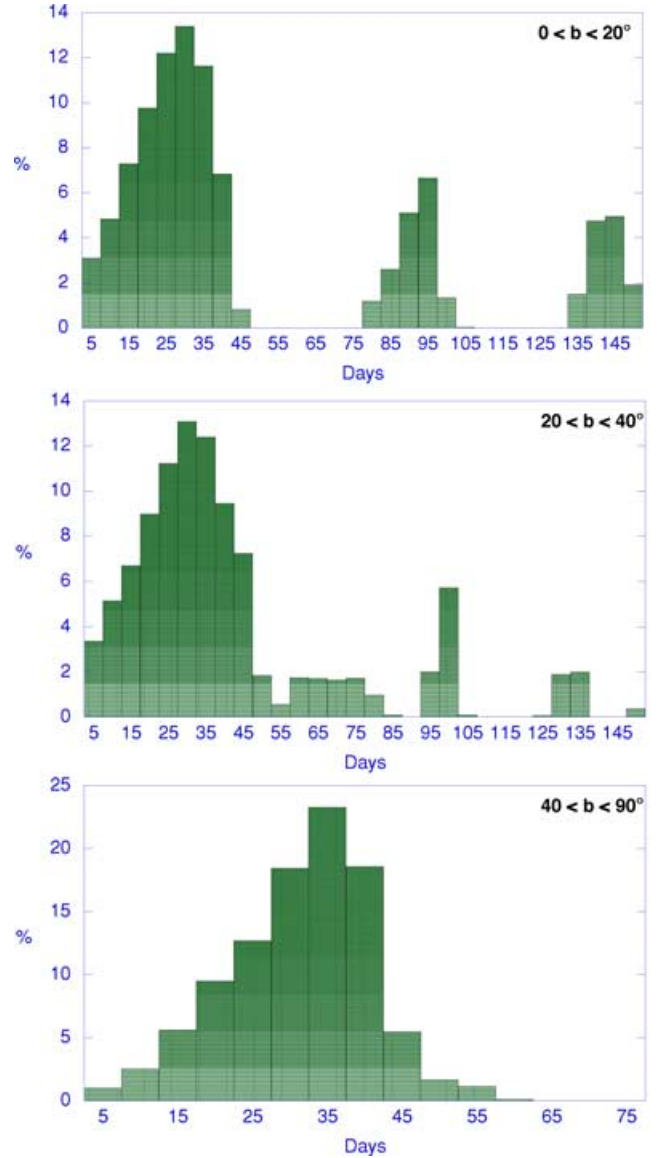
**Figure 5.** Frequency distribution of the intervals between consecutive observations performed at the same epoch in the three photometric FOVs. This indicates that the most common interval between paired observations is the 6 h of the scan period. The lower panel is a close-up view of the overall distribution shown in the upper panel.

The distribution of the short intervals when observations in the three photometric fields are combined is shown in Fig. 5. The most frequent interval is 0.25 d, corresponding to one spin period between two successive observations in the MBP and accounts for 50 per cent of the intervals. The origin of the other intervals can be read in Table 1, by adding in some instances one or two revolutions of 0.25 d.

### 3.2 Long-term succession

The short-term succession within an epoch is essentially driven by the relative position of the fields and the spin period, which control the time needed to rotate from one field to the next. The transverse motion that leads to the end of a sequence and later to the beginning of a new one is due to the slight displacement of the spin axis on the precession cone of  $\sim 4^\circ$  per day. Therefore, the return of an epoch of observation is primarily determined by the precession motion and the yearly solar motion on the ecliptic.

We have seen in the previous section that the average length of a sequence in the astrometric fields is only very close to 2. With a total of  $\sim 80$  observations (Fig. 4), this gives about 40 epochs distributed over the 1800 d of the mission, or a typical interval between epochs of 45 d. If we take the combined observations in the two photometers instead, we have 280 observations and an average sequence of 6.5 observations, that is, a return again every 45 d. However, we have seen that the total number of observations per star is also a function of the ecliptic latitude as any other effect tied to the precessional



**Figure 6.** Frequency distribution (as a function of the ecliptic latitude) of the intervals of time between epochs of observations in the three photometric FOVs. The recurrence is primarily driven by a combination of the precessional motion of *Gaia* spin axis and the yearly solar motion.

motion. The actual number as a function of the latitude can only be obtained through a simulation with a fairly large number of stars.

The frequency distribution for the ecliptic region, the intermediate latitude and the extended polar cap is shown in Fig. 6. These distributions are very informative, and indeed more relevant for the purpose of variable analysis than a weakly scattered distribution about the average interval would be. Below  $b = 40^\circ$ , there is quite an extensive coverage between 5 and 150 d, with two holes in the ranges 50–80 d and 100–130 d. At higher latitude, the scatter is less pronounced, with a smaller period between two epochs, meaning more observations. Clearly, the period recovery of variable stars will depend, not strongly indeed, on the ecliptic latitude just because of the time-sampling. The evaluation of this feature is the main object of the following sections. A similar conclusion applies to the periods of spectroscopic binaries determined from the radial velocity measurements as found by Pourbaix & Jancart (2003).

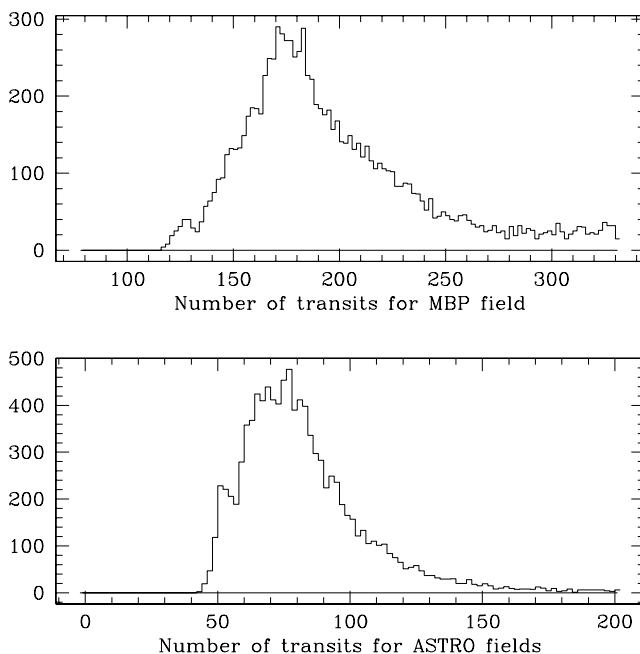
## 4 SIMULATIONS

### 4.1 The time-sampling

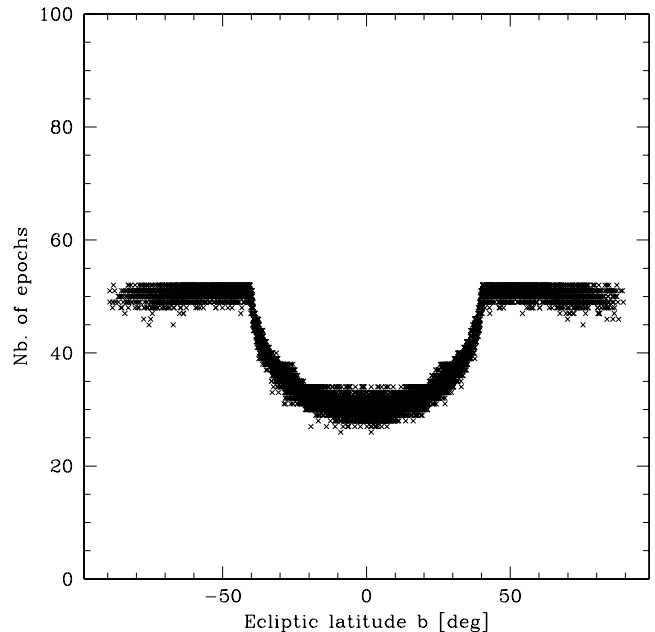
As said earlier, a catalogue of 10 000 uniformly distributed directions has been produced. From the position and size of the different FOVs and the scanning law it was possible to sample the 1800 d of the mission and determine at any time which regions of the sky are observable by each instrument. When a star lies in a field, one then computes accurately the date of the crossing of the central line of the field, together with its ordinate and transverse velocity and the index of the field. Many refinements are used to speed up the run-time by eliminating quickly many of the stars not observable within  $\sim 1$  d. The output is finally re-organized to produce chronological time-series for each of the 10 000 stars. This gives in the end a set of about 3 million observations scattered over 5 yr. This set of realistic observing sequences is then the basis of the photometric simulation and the variability analysis.

The distribution of the number of individual observations (an observation at a particular date corresponds to the crossing of one FOV) is shown in Fig. 4 as a function of the ecliptic latitude for the 10 000 sources of the simulation. There is a strong dependence resulting from the scanning law, with mid-latitude region overobserved compared with the mean, while the ecliptic region is observed less frequently than the average. The same pattern is common to the main astrometric field, which includes the BBP filters, and the MBP field. However, due to the larger across-scan extension of the latter, about twice as many observations are made in the latter field. This appears clearly in Fig. 7, which displays the frequency distribution of the number of observations in the same two FOVs. The scatter in each field is significant with variation from 40 to 150 in the astrometric fields and from 100 to 300 in the MBP field.

As important as the number of observations is the number of different epochs at which the signal is sampled. The observations are clustered into small groups of  $\sim 2$ –10 observations within a day,



**Figure 7.** Histograms of the number of measurements for the astrometric (both following and preceding) and for the MBP field. The extended scatter towards the largest values comes from the observations around  $40^\circ$  of ecliptic latitude.



**Figure 8.** Number of clustered groups as a function of ecliptic latitude  $b$ . The pattern is much simpler than that for the total number of transits (Fig. 4), with only two regimes: low and high ecliptic latitude.

before a new cluster shows itself after several weeks. The number of epochs based on the same simulation plotted in Fig. 8 shows that there are essentially two regimes: the ecliptic zone with  $\pm 30^\circ$  with about 30 observational epochs over the 5 yr and the region of upper ecliptic latitude, above  $40^\circ$ , with  $\sim 50$  different epochs. This feature impacts significantly on the variability analysis probably more than the total number of observations, since it is directly related to the ability to recover the characteristics of periodic phenomena. Even after a normalization over the signal-to-noise (S/N) ratio values, there should remain a systematic effect with the ecliptic latitude.

### 4.2 The photometric signal

For each star, one generates the photometric variability modelled by a simple sinusoidal signal with one frequency. More complex light curves have been tested with several harmonics representing non-sinusoidal oscillations of pulsating variables or eclipsing binaries without affecting the main conclusions of this study, although the case of the eclipsing binaries is the source of specific problems. Therefore, our baseline simulated signal has the form

$$s(t) = S_0 + A \sin(2\pi\nu t + \phi) + \text{noise},$$

where  $A$  is the amplitude of the variability,  $\nu$  the frequency and  $\phi$  the phase. One should note that variable stars are usually characterized by the amplitude of variability, that is, by the peak-to-peak amplitude  $2A$ . The noise added to the deterministic signal is a Gaussian random variable of zero mean with a standard deviation determined by the S/N ratio of each experiment. Although we have also experimented with multiperiodic signals, the results presented here refer to purely periodic variables.

All the simulations are scaled by the S/N ratio (defined here as  $A/\text{noise}$ ). This is a convenient parameter, which makes our results independent of the star magnitude. However, as a result of the systematic variation of the number of observations with the ecliptic latitude, the dominant effect in the variability analysis beyond the



S/N ratio is the ecliptic latitude. In order to avoid this bias, we have used a normalized definition of the S/N ratio by dividing the single-observation noise in the S/N ratio by  $n^{1/2}$ . More precisely we have,

$$\chi = S/N \times \sqrt{n/80}, \quad (2)$$

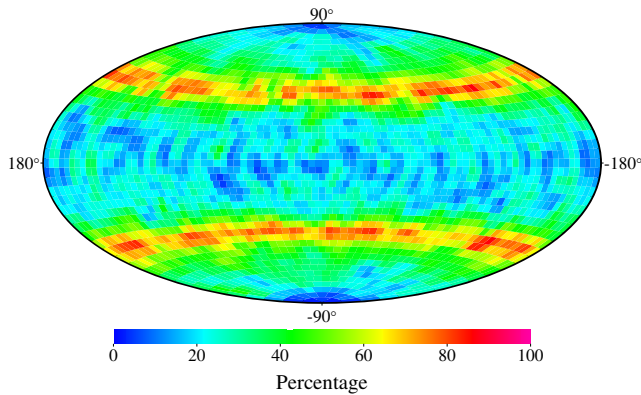
so that  $\chi \sim 1$  for an average observation with  $S/N = 1$ , while at  $b \sim 40^\circ$ ,  $\chi = 1$  is equivalent to  $S/N \sim 0.65$ . Therefore, any remaining effect with the ecliptic latitude will arise more from the time-sampling than from the systematic difference in the number of observations.

Simulated variables are generated for a grid of frequencies and  $\chi$ . For each point of this grid, and for all the sources a period search is performed on the signal, without any a priori information, using primarily an optimized version of the Lomb–Scargle algorithm (Press et al. 1992) or a frequency analysis based on orthogonal decomposition developed by one of us (FM). The rate of success in the period retrieval for every grid point is the main output of the program. Regarding the size of the computation, we have sampled about 10 periods (from the slowest Miras to the shortest pulsating variable with periods of few hours) and between three and five values of the normalized S/N ratio. Each of the 10 000 stars (or 5000 for the runs limited to one hemisphere) has been used five times, with independent random noise on the observations. This gives in total 50 000 period searches for every grid point. In total the number of computed ‘periodograms’ reaches nearly  $1.4 \times 10^6$ .

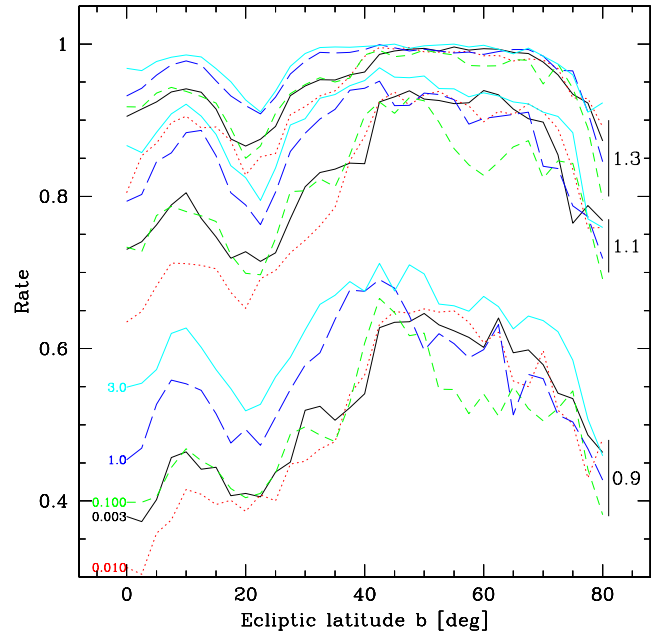
## 5 RESULTS

### 5.1 Rate of detection

Results consist primarily of correct rate of detection as a function of the period, the S/N ratio and the position on the sky, as a result of the inhomogeneous time-sampling of *Gaia*. Fig. 9 gives for a single grid point ( $P = 0.2$  d and the unnormalized  $S/N = 0.75$ ) the sky distribution of the correct rate of period retrieval. The main pattern reproduces the scatter of the number of observations with the ecliptic latitude, which has been subsequently eliminated with the normalized S/N ratio. In the range of latitude around  $40^\circ$  the rate is close to 80 per cent, and between 20 and 50 per cent it is outside



**Figure 9.** Rate of correct detections for  $S/N = 0.75$  and a single period of 0.2 d, using only the astrometric fields. We are here in a range of small S/N ratios. For certain regions, the rate reaches nearly 0 per cent and for other regions where the number of measurements is high, it reaches nearly 100 per cent. The small effects in longitude follow from the non-uniform coverage observation density as a function of longitude over 5 yr.



**Figure 10.** Rate of correct detection for normalized S/N ratios,  $\chi$  (written on the right), and frequencies (cycles per day) (written on the left for the normalized S/N ratio of 0.9, the same colour/grey-scale and line type are applied for the higher S/N ratios).

this area. The rate falls off very quickly with lower S/N ratio, but on the other hand grows quickly to 100 per cent for  $S/N > 1.5$ .

Since there are only small secondary modulations with the ecliptic longitude, one can perform one-dimensional analyses by looking at the dependence with the ecliptic latitude for various grid points. Results are shown in Fig. 10. There are three sets of curves for  $\chi = 0.9, 1.1, 1.3$  and different periods (8 h, 1, 10, 100 and 330 d).

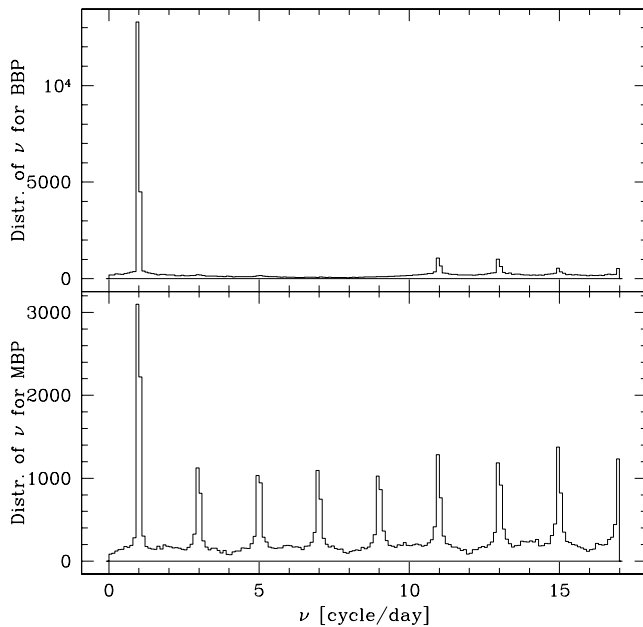
(i) It is clear that, despite the normalization of the S/N ratio, the effect with the ecliptic latitude has not fully vanished, indicating an influence of the number of epochs (the number of clusters of observations separated by a few weeks as shown in Fig. 8). There is a systematic loss in the period recovery around  $b \simeq 20^\circ$ , which is not easily understood at the moment.

(ii) The rate of correct recovery grows quickly to 100 per cent, with normalized  $S/N \geq 1.5$ .

(iii) Although the scatter is large, it seems that the shorter periods ( $< 1$  d) are better recovered than the larger periods. However, the difference is not conspicuous, and this comes as a surprising (and welcome) result, as it was thought that very short periods would be very difficult to recover.

### 5.2 Aliasing and false periods

With irregular sampling, and within the range of periods analysed here, one should not expect too much aliasing (Eyser & Bartholdi 1999), since the somewhat randomness of the sampling limits the building up of coherent signature at regularly spaced frequencies. However, at low S/N ratio levels the recovery rate is low, and the search program may end up in a normal exit with no periodic signal found (the favourable outcome) or with the identification of a period (possibly wrong), without warning. It is then of interest to look at what kind of periods are found as a function of the true period. This is illustrated in Fig. 11 for the BBP (same field as the astrometric



**Figure 11.** Example of terminal frequencies for an initial signal of  $S/N = 0.75$  and frequency  $\nu = 1$ . It reflects the pattern of individual periodogram, which bears the signature of the spectral window.

detector) (top) and the MBP fields (bottom). The simulated variable has a period of 1 d and  $S/N = 0.75$  in this example. In the BBP, when the period is not recovered (about half of the cases tested) the final erroneous frequency can be almost anything, with a higher probability to frequencies of 11 and 13 cycles  $d^{-1}$ , coming from the convolution with the spectrum of the observing window. Therefore, unless further assumptions are made, there is no way to decide from the output of the period search whether the period found is correct, although a significance level is easily produced and could be used as a warning, but not as a measurement of the uncertainty of the period. In the case of the MBP, aliased results are clearly visible, regularly spaced with  $\delta\nu = 2$  cycles  $d^{-1}$ , corresponding to half of the satellite spin frequency (period of 6 h). This comes from successive observations regularly sampled for every revolution within a cluster of observations, responsible for the aliasing and leading to a wrong determination of the period. This again cannot be avoided with the present design of the instrument, but fortunately limited to those observations with low  $S/N$  ratio.

## 6 CONCLUSION

In this paper, we have investigated the retrieval of periods of variable stars from observations to be carried out during the *Gaia* mission. We have found that with adequate frequency analysis algorithms one can recover the period of regular variable stars even with a  $S/N$  ratio of around unity. It is also shown that the very irregular time-sampling expected from *Gaia* allows to determine frequencies much higher than the upper frequency estimated from a straightforward evaluation of the Nyquist boundary (cf. Eyer & Bartholdi 1999), and with virtually no aliasing. These two features are very good news for the preparation, and subsequent scientific exploitation, of the mission.

However, from this investigation we have noted that the frequency analysis without initial assumption on the frequency range may be very demanding in terms of computer resources, which are

hardly compatible with the goal of analysing from the scratch ( $\sim 10^8$  sources). Therefore, a major effort is needed during this preparatory phase to develop optimized software, to estimate quickly and efficiently possible periods or ranges of periods, before a thorough and accurate search can be launched on a well-bounded frequency range. An estimate of the level of significance of each peak height should also be a built-in feature, alongside the possibility of searching for harmonics or a second independent period.

## ACKNOWLEDGMENTS

We would like to thank Michel Grenon for interesting discussions and Floor van Leeuwen for his valuable comments on the article.

## REFERENCES

- Eyer L., Bartholdi P., 1999, *A&AS*, 135, 1
- Eyer L., Cuypers J., 2000, in Szabados L., Kurtz D. W., eds, *ASP Conf. Ser. Vol. 203, Proc. IAU Colloq. 176, The Impact of Large-Scale Surveys on Pulsating Star Research*. Astron. Soc. Pac., San Francisco, p. 71
- Eyer L., Grenon M., 1997, *ESA SP-402*, 467
- Jordi C., Carrasco J. M., Figueras F., Torra J., 2003, in Munari U., ed., *ASP Conf. Ser. Vol. 298, Gaia Spectroscopy, Science and Technology*. Astron. Soc. Pac., San Francisco, p. 209
- Mignard F., 2001, *A Practical Scanning Law for Gaia Simulations*, Technical Note GAIA\_FM\_010
- Perryman M. A. C. et al., 2001, *A&A*, 369, 339
- Pourbaix D., Jancart S., 2003, in Munari U., ed., *ASP Conf. Ser. Vol. 298, Gaia Spectroscopy, Science and Technology*. Astron. Soc. Pac., San Francisco, p. 345
- Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, *Numerical Recipes in FORTRAN*. Cambridge Univ. Press, Cambridge

## APPENDIX A: DERIVATION OF EQUATION (1)

Let a star be observed in one field and crossing the field in the along-scan direction at the ordinate  $y$ . The field height is  $h$  and  $0 < y < h$  during the transit. This crossing may happen at any ordinate with a uniform probability, so that  $P(Y < y) = y/h$ . After a time-interval of  $\Delta t$  (one of the possible intervals between the preceding and following fields listed in Table 1) the rotation of the satellite has brought the line of sight of the second field in the same direction and the star may be observed again, provided its transverse motion during  $\Delta t$  has not been too large. Mathematically, the star cannot be observed again if  $y + V_t \Delta t > h$ , meaning that its new ordinate is larger than the field height (we have restricted to the case  $V_t > 0$  to simplify the discussion, but this does not alter the final result for symmetry reasons). Therefore, the probability that the star is not observed in the second field when the transverse velocity is  $V_t$  is

$$P((X = 1) | V_t) = \begin{cases} 1 & \text{if } V_t \Delta t \geq h \\ P(Y > h - V_t \Delta t) = \frac{V_t \Delta t}{h} & \text{if } V_t \Delta t \leq h. \end{cases}$$

In the first case, the transverse velocity is large enough for the transverse motion during  $\Delta t$  to be greater than the field width. In the second case, there exists a region in the initial FOV where even after the time  $\Delta t$  the ordinate of the star is still less than the field width. The above probabilities are conditioned on the transverse velocity  $V_t$ , and one must now introduce its random distribution. Assuming that the star can be anywhere on the scan circle with a uniform



probability, one has  $V_t = V_{\max} \sin(\phi)$ , where the maximum velocity  $V_{\max} = 0.170h^{-1}$  and  $\phi$  is uniformly distributed between 0 and  $2\pi$ . The final probability is just the average of the conditional probabilities over the distribution of the transverse velocity (restricted here to the positive values). The case when  $V_{\max} \Delta t/h \leq 1$  is very simple is

$$P(X = 1) = \frac{2}{\pi} \frac{V_{\max} \Delta t}{h} \int_0^{\pi/2} \sin \phi \, d\phi = \frac{2}{\pi} \frac{V_{\max} \Delta t}{h}.$$

In the second case, the integration extends up to  $\phi_m$ , such that  $V_{\max} \Delta t \sin \phi_m = h$ , or  $\phi_m = \arcsin(h/V_{\max} \Delta t)$ , and beyond this angle the conditional probability is 1. This yields for  $V_{\max} \Delta t/h \geq 1$ ,

$$P(X = 1) = \frac{2}{\pi} \frac{V_{\max} \Delta t}{h} \int_0^{\phi_m} \sin \phi \, d\phi + \frac{2}{\pi} \int_{\phi_m}^{\pi/2} d\phi,$$

giving with  $u = h/(V_{\max} \Delta t)$  (hence  $u < 1$  here),

$$P(X = 1) = \frac{2}{\pi} \frac{1}{u} [1 - (1 - u^2)^{1/2}] + \left[ 1 - \frac{2}{\pi} \arcsin(u) \right],$$

which is the same as equation (1).

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.